

Assignment 3-Sequence-Structure Relationships

Group F

1. Jack Goon
2. Reyna Khalilieh
3. Jacqueline Ling
4. Emily de Paz
5. Jie Tang

Protein: Ribose-5-Phosphate Isomerase

• **PROTEIN NAME AND ORGANISM:** Ribose-5-Phosphate Isomerase in E. Coli

• SEQUENCES

FIVE ORTHOLOGS FOR MULTI-SEQUENCE ALIGNMENT

#1 Organism *Pyrococcus horikoshii* Accession number Q50083

2 Organism: *Streptococcus mutans serotype c* Accession number Q8DTT9

#3 Organism *Thermus thermophilus* Accession number Q72J47

4 Organism *homo sapiens* Accession number P49247

#5 Organism *Bartonella henselae* Accession number Q6G3V6

• MULTISEQUENCE ALIGNMENT ANNOTATIONS

SEQUENCE ALIGNMENT AND ANNOTATIONS-Alignments of regions that are the most diverse 70aa region and annotations (see above)

Residues 136-195

```

13  14      15      16      17      18      19
67890123456789012345678901234567890123456789012345 ID
IPMARSAVARQLVKLGGRPEYRQGVVTDNGNVILDVHGMEILDPIAMENAINAIPGVVTV ECOLI
.SSBSSBB.SSBSSSSSS.S.SSS.S.S.S..BBSBSS.SBSS.SSB.SSBBSBBBBBBB EXP
.....H.....P.....PP..PHH.P.....H.P...H...H....CHH.. CHAR
..LSLLS..L.L.L.L.SSLL.L.L.LLSSSSSLLLS.LLSLL.L.L.L...L..L.S.L.S SIZ
.....ss... FUNC

```

SEQUENCE ALIGNMENT AND ANNOTATIONS-Alignments of regions that are the least diverse 70aa region and annotations (see above)

Residues 81-150

```

8      9      10      11      12      13      14
12345678901234567890123456789012345678901234567890 ID
DGADEINGHMOMIKGGGAALTREKIIASVAEKFICIAADASKQVDILGKFPLPEVIPMARSAVARQLVKL Seq
BBBSSBSSS..BBBBSSS.BBS.BBBB.BBSSBBBBB.SSSSBSSB.S..BBBBBBS.BSSBSSS.BSSB EXP
PCHPPH...H..HPC.C.HH..PPHH.....HHHHP.PP....HC..HCHPH.....H... CHAR
SSSSLSS.LSLLLS.SSSL.LLLLL...S..LL.LSSL.LLS..LS.L.L.LLL.LLSLLS..L.L.L. SIZ
C..C.....C.....css.....b..b..... FUNC

```

• IMAGES

-BACKBONE REPRESENTATION-Show a backbone representation of a single chain the crystal structure of your polypeptide. Indicate the most-conserved 70aa segment and the least-conserved 70aa segment that you used in your MSAs above.

Red = conserved (least diverse)

Cyan = most diverse

Yellow = overlap between segments (this area is diverse and has no conserved residues. It was only added to the "conserved" region so that the conserved region would meet the 70 aa requirement. Probably best grouped together with the cyan section.)

Note: The image shows the homodimer form



-CONSERVED SITE-Pick a residue in the stretch of the 70 aa of most-conserved residues. -----What is the residue # and type:

92 and methionine

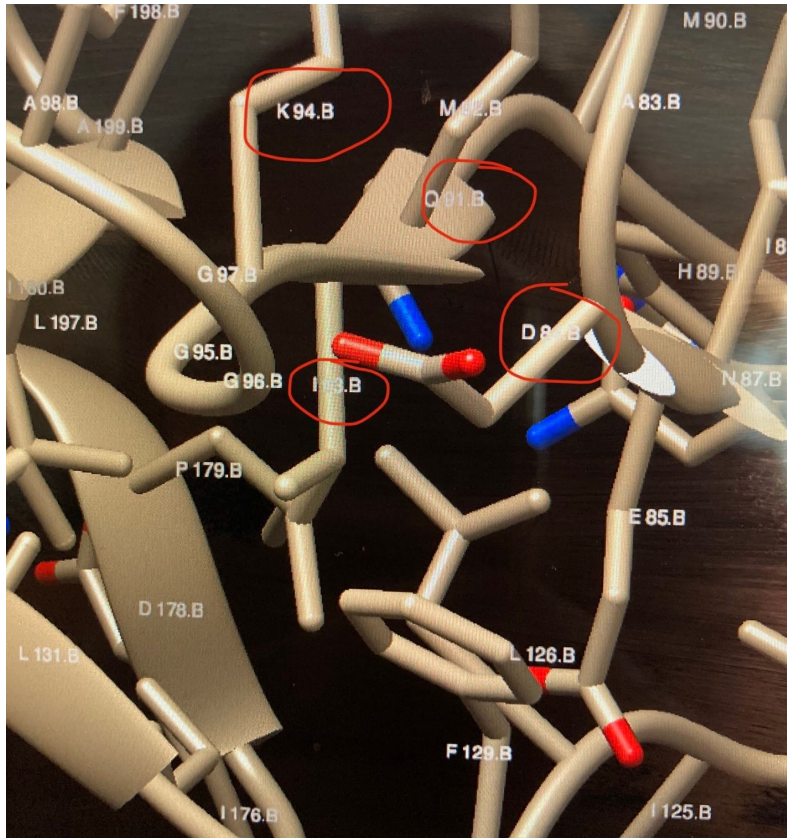
-Find the four closest (spatially) residues to this residue. What are they (residue type and #)?

#94 and lysine #93 and isoleucine #84 and aspartic acid #91 and glutamine

-Comment about the level of conservation of these residues.

Residue 94, 93, and 84 are highly conserved residues. Residue 91 is not a conserved residue.

-Show a close up image of the residue and the environment around it. Indicate the four residues. (use the drawing feature of MS Word)



• QUESTIONS

- What are the characteristics of the most completely conserved residues in terms of

exposure: The most conserved residues were largely buried, with only a few exceptions like 163 and 170.

character: There was decent diversity of character in the conserved residues. Overall, polar residues were more often completely conserved. This is probably to retain salt bridges and catalytic sites, which are essential for structure and function, respectively.

size: For the whole sequence, there are generally more large amino acids than small ones. This is also true of the conserved residues. There was not a clear connection between conservation and a specific size. Some residues were small and conserved, like two of the three catalytic residues. There were also conserved large residues like I93. Since this residue is large and buried, it would be hard to replace it with anything else without causing strain. This is one explanation for why a large residue might be conserved.

- What are the average differences in the below between the most and least conserved stretches?

degree of burial - The most conserved stretch was at least 48% buried while the least conserved stretch was just 30% buried. On average, the more conserved residues were buried more. There were a good number of ambiguous residues. With further analysis (probably requiring actual ASA calculations), we could get a more accurate look at degree of burial because we could quantify partially-buried residues.

characteristics - The sequence consisted of residues with a fair amount of conserved characteristics whether or not the residue type was conserved. 21% of the most diverse region consisted of residues with conserved characteristics whereas 45.7% of the least diverse region had conserved characteristics. Most of the conserved characteristics were hydrophobic with few polar and conformational residues.